

Semantic Learning Machine Improves the CNN-Based Detection of Prostate Cancer in Non-Contrast-Enhanced MRI

Paulo Lapa

NOVA IMS, Universidade Nova de Lisboa
Lisboa, Portugal
plapa@novaims.unl.pt

Leonardo Rundo

Department of Radiology, University of Cambridge
Cambridge, UK
lr495@cam.ac.uk

Ivo Gonçalves

INESC Coimbra, DEEC, University of Coimbra
Coimbra, Portugal
ivogoncalves77@gmail.com

Mauro Castelli

NOVA IMS, Universidade Nova de Lisboa
Lisboa, Portugal
mcastelli@novaims.unl.pt

ABSTRACT

Considering that Prostate Cancer (PCa) is the most frequently diagnosed tumor in Western men, considerable attention has been devoted in computer-assisted PCa detection approaches. However, this task still represents an open research question. In the clinical practice, multiparametric Magnetic Resonance Imaging (MRI) is becoming the most used modality, aiming at defining biomarkers for PCa. In the latest years, deep learning techniques have boosted the performance in prostate MR image analysis and classification. This work explores the use of the Semantic Learning Machine (SLM) neuroevolution algorithm to replace the backpropagation algorithm commonly used in the last fully-connected layers of Convolutional Neural Networks (CNNs). We analyzed the non-contrast-enhanced multispectral MRI sequences included in the PROSTATEx dataset, namely: T2-weighted, Proton Density weighted, Diffusion Weighted Imaging. The experimental results show that the SLM significantly outperforms XmasNet, a state-of-the-art CNN. In particular, with respect to XmasNet, the SLM achieved higher classification accuracy (without neither pre-training the underlying CNN nor relying on the backpropagation) as well as a speed-up of one order of magnitude.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks; Bio-inspired approaches; Search methodologies; Learning settings;**

KEYWORDS

Prostate cancer, Non-contrast-enhanced MRI, Semantic Learning Machine, Neuroevolution, Convolutional Neural Networks

ACM Reference Format:

Paulo Lapa, Ivo Gonçalves, Leonardo Rundo, and Mauro Castelli. 2019. Semantic Learning Machine Improves the CNN-Based Detection of Prostate Cancer in Non-Contrast-Enhanced MRI. In *Proceedings of the Genetic and*

Evolutionary Computation Conference 2019 (GECCO '19). ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3319619.3326864>

1 INTRODUCTION

According to the American Cancer Society, in 2019 it is expected that the Prostate Cancer (PCa) is the most common malignant tumor in Western men [43]. Nevertheless, prostate image analysis is still a compelling challenge in every diagnostic and therapeutic phase of the PCa pathology. Nowadays, high-resolution multiparametric Magnetic Resonance Imaging (MRI) is acquiring clinical and scientific interest, also enabling quantitative studies [46].

As a matter of fact, in addition to conventional anatomical information conveyed by T1-weighted (T1w), T2-weighted (T2w) or Proton Density-weighted (PDw) MRI protocols, complementary and valuable functional information regarding the tumor can be acquired by means of specific MRI sequences [25], such as: Dynamic Contrast Enhanced (DCE) [9] for evaluating the perfusion and vascularity in the tumor micro-environment; Diffusion Weighted Imaging (DWI) to estimate the motion of water molecules within tissue voxels [17] as well as to quantitatively measure the resistance of water molecule diffusion by using the Apparent Diffusion Coefficient (ADC) maps; Magnetic Resonance Spectroscopic Imaging (MRSI) to quantify the metabolic activity [31]. Unfortunately, this potential is strongly reduced by the difficulties related to the development of accurate detection and segmentation methods that can effectively support experienced radiologists in their decision-making tasks [18], overcoming the limitations imposed by the fully manual procedures. In practice, a standardized interpretation of multiparametric MRI might be affected by a significant inter-observer variability [36], often overwhelming the analytic capabilities of radiologists in their diagnostic workflows. Therefore, computer-assisted MR image analysis approaches are mandatory for improving results repeatability in large-scale clinical studies [37].

Despite the vast amount of studies aimed at improving the performance of CNNs for classification tasks, the majority of these contributions has been focused on the earlier layers of the neural network, meaning everything up to the fully-connected layers [44]. Thus, fully-connected layers represent the common choice for building the classifier that, after receiving in input the features obtained from the previous layers, produces the prediction about the class label. While this choice produced excellent performance over several problems [45], it is fundamental to perform additional investigation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

GECCO '19, July 13–17, 2019, Prague, Czech Republic

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6748-6/19/07...\$15.00

<https://doi.org/10.1145/3319619.3326864>

in this area. In fact, according to the no-free-lunch theorem [51], no classifier can be said to be superior to all other classifiers across all the problem instances. Thus, the idea explored in this work is to improve the performance of the final model by using a network created by a neuroevolution algorithm on high-level image features. More in detail, the focus of this paper is not directly on replacing the fully-connected layers in the learning process, but rather improving it by means of a two-step procedure: (1) **Regular CNN Training Process** in which the usual training phase of a given CNN architecture is performed to make convolutional layers learn image features from image inputs, and (2) **Improvement of the Training Process** in which the fully-connected layers are replaced by a network obtained by a newly proposed neuroevolution algorithm and the training process exploits the high-level image features (in terms of meaningful representation learning [48]), yielded by the convolutional layers in the first step.

To achieve the challenging objective previously described, this work relies on a recently defined neuroevolution algorithm called Semantic Learning Machine (SLM) [12]. As stated in the original contribution, SLM is able to induce a unimodal error landscape in any supervised learning problem where the error is measured as a distance to the known targets. Thus, the algorithm is able to construct the artificial neural network architecture during the training phase, without being trapped in local optima. This is a fundamental advantage with respect to the use of the traditional multi-layer perceptron architecture trained with the backpropagation algorithm, and it is the main reason for selecting the SLM algorithm in this work. We expect that the SLM is able to outperform the fully-connected neural network as already shown on a vast set of benchmark problems [19]. The workflow, from the input MRI series to the training and validation of the neural models, is shown in Figure 1. The various processing steps are explained in the subsequent sections. In addition, a schematic representation of the intermediate data creation (performed by the CNN model) and the subsequent SLM-based classification are reported in Figure 2.

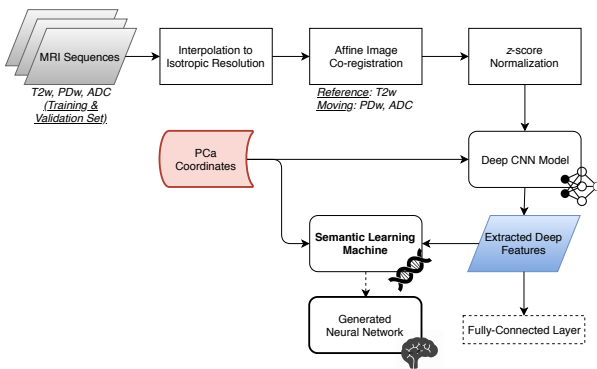


Figure 1: Workflow of the proposed neuroevolution approach based on the SLM

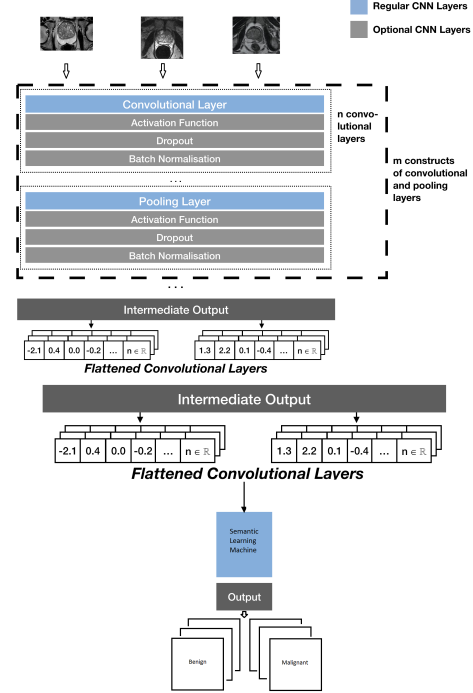


Figure 2: Intermediate Data Creation (CNN) and subsequent Classification (SLM)

The paper is organized as follows: Section 2 outlines the state-of-the-art of the CNNs in classification tasks, with particular interest to prostate MR image analysis; Section 3 introduces the SLM algorithm, highlighting its properties; Section 4 describes the experimental settings and presents the dataset taken into account, as well as the selected parameters; Section 5 presents and discusses the obtained results. Finally, Section 6 concludes the paper and suggests future research directions.

2 BACKGROUND

This section presents existing studies where CNN-based models were developed for addressing the PCa classification problem posed by the PROSTATEx Challenge [2]. This is the challenge that was considered in the experimental campaign described in Section 4. In particular, after presenting the most important works in this area, the CNN-based model taken into account in this paper, called XmasNet [29], will be described. Its salient properties will be recalled, and the motivations for considering it in this study will be outlined.

In the latest years, deep learning techniques are achieving outstanding results in prostate segmentation tasks by using deep feature learning combined with shape models [16] or location-prior maps [48]. CNNs were employed with patch-based ensemble learning [20] or dense prediction schemes [32] for whole prostate gland segmentation. With respect to prostate zonal segmentation, Clark *et al.* [8] detected the DWI MR slices with prostate relying on Visual Geometry Group (VGG) net [44], and then sequentially segmented

the whole gland and the transition zone by using U-Net [35]. In the case of PCa detection and diagnosis on multiparametric MRI, end-to-end deep learning models were proposed and tested on the PROSTATEx dataset [2]. In [29], a VGG-based architecture [44], called XmasNet, outperformed a machine learning model based on an ensemble of decision trees relying on hand-crafted features. Wang *et al.* [50] combined a sub-network for automated prostate detection and multispectral image registration, with a dual-path CNN for clinically significant (CS) PCa detection. This joint approach was compared against 2D U-Net [35] and 3D DeepMedic [22], obtaining the highest classification accuracy. However, the automated analysis of radiomics features, based on traditional machine learning models [24], is gaining interest in the assessment of the heterogeneity in PCa [46].

Even though several CNN architectures were used in medical imaging, they have to be specifically adapted for the medical imaging application at hand. In this work, we considered the XmasNet architecture [29] as a deep feature extractor and for the prediction between benign and malignant PCa.

Indeed, XmasNet [29] was developed *ad hoc* for the PROSTATEx 2017 challenge. Despite the simple architecture (four convolutional and two fully-connected layers), it achieved state-of-the-art performance in the challenge. Thanks to the detailed description provided by the authors in the paper [29], we carefully reproduced this architecture in our implementation. As a matter of fact, along with classic Evolutionary Computation techniques for medical image enhancement [41], segmentation [4] and quantification [39], the aim of our exploratory study is to show the potential of neuroevolution algorithms in clinical applications exploiting deep learning.

The XmasNet architecture is based on VGG net [44], which was the first truly deep network, achieving the first and second place in the ImageNet 2014 competition for the localization and classification challenges, respectively. For instance, VGG-16 employed 16 layers: 13 convolutional and three fully-connected. The main achievement of this network was the considerable performance improvement that was obtained by using 16-19 weight layers, while revealing good generalization capabilities on other datasets [44]. For these motivations, we selected XmasNet as an end-to-end trainable deep model for the PROSTATEx dataset. XmasNet allows for efficient training and we aim at significantly improving its performance by means of the SLM-based neural network classifier.

3 SEMANTIC LEARNING MACHINE

The SLM neuroevolution algorithm [12] was created by deriving a mutation operator proposed in Geometric Semantic Genetic Programming (GSGP) [33] to be applicable to the construction of Neural Networks (NNs). In GSGP, Moraglio *et al.* [33] showed that any supervised learning problem where the error is measured as a distance to the known targets has a unimodal error landscape that can be effectively explored by constructing specific variation operators. These operators are known as geometric semantic operators. In this context, the term semantics is used to refer to the outputs of any supervised learning model (e.g., an NN) over a set of data instances. The reasoning behind these geometric semantic operators can be used to create equivalent operators for other representations or computational models. The SLM was created by defining

a geometric semantic mutation for NNs. This allows SLM to explore unimodal error landscapes and to effectively deal with the space of NNs. Given that these error landscapes are unimodal, no local optima exist, i.e., with the exception of the global optimum, every point in the search space has at least one neighbor with better fitness, and that neighbor is reachable through the application of the variation operators. This translates into the fact that the SLM, as well as GSGP, can simply follow a hill climbing strategy to effectively advance the search.

In essence, the SLM neuroevolution algorithm is a geometric semantic hill climber for NNs. Without local optima, the search can be focused around the current best NN without incurring in any particular disadvantage. The SLM procedure can be summarized in the following steps:

- (1) Generate N initial random NNs
- (2) Choose the best NN (B) from the initial random NNs, according to the selected performance criterion
- (3) Repeat the following steps until a given stopping criterion is met:
 - (a) Apply the geometric semantic mutation to the current best (B) N times to generate N new NNs (known as children or neighbors)
 - (b) Update B as being the NN with the best performance according to the selected criterion, considering the current B and the N newly generated NNs
- (4) Return B as the best performing NN according to the selected performance criterion

The initial random NNs can be generated without any particular restriction. They can have any number of layers and neurons, with any activation function, while the weights in the connections between the neurons can be freely selected. As it is common in neuroevolution algorithms, the SLM does not rely on backpropagation to adjust the weights of the NNs. The crucial aspect is the definition of the geometric semantic mutation for NNs. This mutation essentially works by adding new neurons to the existing hidden layers. Each new neuron can select from which neurons it receives incoming connections. This means that the NN does not need to be fully-connected and its sparseness can be controlled by defining how many incoming connections each new neuron will receive, given the set of all possible incoming connections. The weights of each connection can be freely selected as in the initialization step. A fundamental aspect of this geometric semantic mutation is that the new neurons do not feed their computations to existing neurons, with the exception of the output neuron. The weights of connections from the new neurons in the last hidden layer to the output neuron are defined by the learning step. The learning step can be computed optimally by means of the Moore-Penrose pseudoinverse (similarly to the case of GSGP [11, 13, 34]), or it can be defined as a parameter to be tuned. For full details regarding the SLM as its possibilities the reader is referred to Gonçalves *et al.* [12].

An additional interesting characteristic of geometric semantic methods, such as SLM and GSGP, is the feasibility of using the information gathered from the semantic neighborhood (the set of new models generated by the mutation) to decide when to stop the search. This can be achieved by using the Semantic Stopping

Criteria (SSC) proposed by Gonçalves et al. [14]: the Error Deviation Variation (EDV) criterion and the Training Improvement Effectiveness (TIE) criterion. EDV stops the search when a considerable majority of the neighbors are improving the training performance at the expense of larger error deviations. TIE stops the search when training error improvements become harder to find within the semantic neighborhood. This can signal that the training error improvements are being forced at the expense of the resulting generalization. These SSC can be used to avoid setting a maximum number of iterations as well as leaving data aside to use as a validation set to decide when to stop. For a detailed description of the SSC, the reader is referred to Gonçalves et al. [14].

4 DATASET AND EXPERIMENTAL SETTINGS

This section describes the data used in the experimental phase (Section 4.1), as well as the parameter setting (Section 4.2). This allows for the reproducibility of the results discussed in Section 5.

4.1 The PROSTATEx Dataset

In this work, we consider the prostate MRI dataset provided by the PROSTATEx Challenge [2], organized by the International Society for Optics and Photonics (SPIE) and supported by the American Association of Physicists in Medicine (AAPM) and the National Cancer Institute (NCI). This challenge aimed at encouraging the development of novel quantitative image analysis methods for the diagnostic classification of CS PCa [7, 28]. The prostate MRI dataset was retrospectively collected and curated by the Radboud University Medical Centre (Nijmegen, The Netherlands) in the Prostate MR Reference Center [27]. All the studies included T2w, PDw, DCE, and DWI MRI sequences, acquired on two different types of Siemens (Siemens Healthineers, Erlangen, Germany) 3T MRI scanners, namely MAGNETOM Trio and Skyra, without using an endorectal coil. For more details, please refer to [2].

The whole MRI dataset used for the PROSTATEx competition consisted of a training dataset of 204 patients (with 330 suspected lesions) and a test set of 140 patients (with 208 suspected lesions). The labels of the CS PCa are encoded by means of the approximate location of the centroid of each lesion, and represent the presence of a clinically significant finding (Gleason score ≥ 7), or not. The performance of our neuroevolution method were evaluated only on the PROSTATEx training set that includes 70 MRI-targeted biopsy-proven CS PCa and 134 non-CS PCa patients. The test data from the PROSTATEx competition were not usable for our purposes since the ground-truth labels (i.e., CS or non-CS) are not provided [2]. It is worth noting that the PROSTATEx challenge uses biopsy points as a ground-truth instead of the corresponding tumor segmentations [20, 47]. By taking into account the available MRI sequences in the PROSTATEx dataset as well as by following the current clinical trend that aims at reducing the use of contrast medium administration [49], in this study we analyzed only the non-contrast-enhanced MRI sequences, namely: T2w, PDw, and ADC series. As reported in [5], good quality results in PCa detection can be achieved just with anatomical and DWI MRI data, without analyzing DCE MRI sequences. In the latest years, along with economic factors regarding DCE MRI, the safety of Gadolinium-based contrast agents has

been debated due to the evidence of possible depositions [21], such as in the brain [15].

Since the MR images were collected under different conditions (e.g., different scanners and acquisition protocols), a pre-processing phase was performed. These steps are graphically outlined in Figure 1. The analyzed MRI series are characterized by different image resolution, depending on the pixel spacing and slice thickness. Therefore, the images were interpolated to an isotropic resolution of 1.0 mm. To ensure that the prostate MRI features are aligned in the same reference space in every sequence, an affine co-registration was performed. In particular, intensity-based co-registration applies a set of transformations on one image (the moving image) against a reference (i.e., fixed) image, to maximize the matching of the pair [40]. The voxel normalization and the image registration were performed by using the Diffusion Imaging Python (DIPy) [10] package, and the Mutual Information metric was considered as the cost function [30]. In all the cases, we considered the T2w sequence as the reference series, while ADC and PDw sequences as the moving series. By doing so, all the MRI sequences are brought onto the same reference system. Thus, an early fusion scheme can be employed for constructing feature vectors by concatenating the corresponding pixels in the three different series [42].

For the classification task at hand, each lesion was extracted by considering a region of 32×32 pixels where the lesion coordinates were used to identify the center of the Region of Interest. After this step, 335 lesions were extracted, including 20.5% of CS PCa. Finally, before feeding the feature vectors to the CNN-based model, the pixel intensities of each series were standardized to zero mean and unitary standard deviation (i.e., z-score normalization), to accelerate the convergence and reducing the chance of getting stuck in local optima during the training phase.

4.2 Parameter Tuning of XmasNet

To estimate the best performing configuration of the considered CNN, a random search was performed. The original data set was divided into three parts: training, validation and test. In particular, 60% was used for training purpose, 20% for validation, and 20% for testing the performance of the model over unseen instances. 30 random configurations were tested and for each configuration the model was trained on the training and validation sets 30 times, by averaging the achieved results. The configuration that returned the lowest average loss (i.e., binary cross entropy) on the validation set was selected as the best model. The best configuration was evaluated 30 times on the unseen test data by taking into account the Area Under Receiver Operating Characteristic (AUROC).

More in detail, the random search took into account the learning rate α as well as the following parameters:

- Stochastic Gradient Descent (SGD) [3]: weight decay β , momentum and use of Nesterov momentum;
- Root Mean Square Propagation (RMSProp): weight decay β and use of the AMSGrad variant;
- Adam [23]: weight decay β .

The best configuration obtained an average test loss of 0.0405 ($\sigma = 0.0077$) and a AUROC of 0.517 ($\sigma = 0.018$). It used RMSPROP with $\alpha = 1 \times 10^{-5}$ and $\beta = 0.999$.

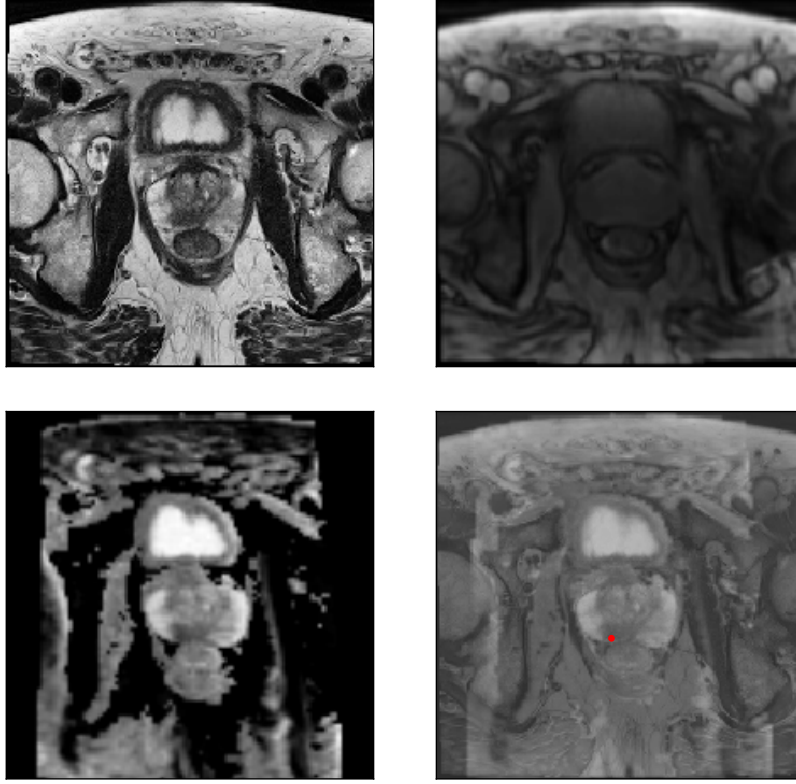


Figure 3: From left to right, up to down: T2w, PDw, and ADC axial MR images extracted from Patient #29 of the dataset, after being interpolated and co-registered. Concatenation of three previous MR images, composing the feature vectors fed to XmasNet. The PCa coordinates are highlighted in red

The different configurations of XmasNet were trained on a single GPU, using Keras [6] with TensorFlow as back-end [1]. The computational platform used was an ASUS X550L laptop, with an NVIDIA 840M GPU having 2GB of RAM and 2.0 GHz of clock frequency. The laptop has a 4th generation Intel Core i7-4510U processor with 4 cores, 2.0 GHz of base frequency, and 8 GB of RAM. SLM was also trained on the same computational platform.

4.3 Parameter Tuning of SLM

The SLM base configuration considered is the following:

- In the initial population each NN is generated with a random number of hidden layers selected between 1 and 5
- In the initial population each NN randomly selects the number of neurons for each hidden layer between 1 and 5
- Each hidden neuron randomly selects its activation function from the following options: logistic, Rectified Linear Unit (ReLU), and tanh
- Each hidden neuron randomly selects the weight of each incoming connection from values in the interval $[-mncw, mncw]$, where $mncw$ represents the maximum neuron connection weight parameter (subject to parameter tuning)

- Each hidden neuron randomly selects the weight of its bias from values in the interval $[-mbw, mbw]$, where mbw represents the maximum bias weight parameter (subject to parameter tuning)
- Every time a new NN is created by the mutation operator, the number of new neurons to be added to each layer is randomly selected between 1 and 3
- The weights of connections from the new neurons in the last hidden layer to the output neuron are selected from values in the interval $[-mls, mls]$, where mls represents the maximum learning step parameter (subject to parameter tuning)

To ensure fairness, the SLM also explores 30 random parameter configurations as XmasNet. The possible parameter configurations and their ranges are the following:

- Maximum learning step (mls): $[0.1, 1]$
- Stopping criterion: EDV or TIE (with equal probability)
- Maximum neuron connection weight ($mncw$): $[0.1, 0.5]$
- Maximum bias weight (mbw): $[0.1, 0.5]$

The best configuration found according to the average validation performance is the following: maximum learning step (mls) of 0.25, use of the EDV stopping criterion, maximum neuron connection weight ($mncw$) of 0.17, and maximum bias weight (mbw) of 0.37.

5 RESULTS

This section analyzes the results achieved in the experimental phase. In particular, the objective is to compare the classification performance achieved by XmasNet against the one obtained with the SLM. Additionally, it is fundamental to understand whether the used neuroevolution algorithm is comparable with XmasNet also in terms of execution time. In fact, the combination of the two factors (classification performance and execution time) is fundamental for applying the SLM over challenging tasks characterized by thousands (or more) images.

To analyze the classification performance, the Area Under the Receiver Operating Characteristic (AUROC) curve was taken into account, considering that the dataset is not balanced. The results of this comparison are reported in Figure 4.

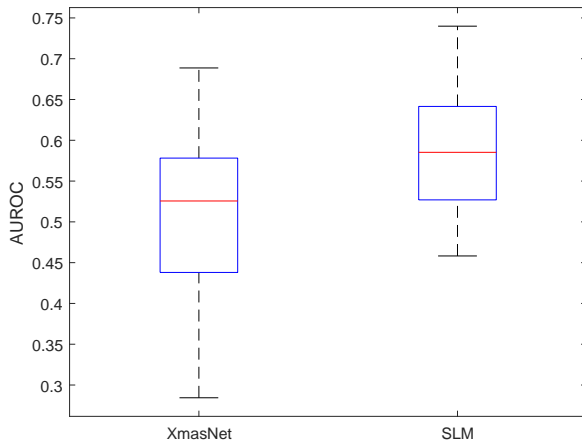


Figure 4: AUROC achieved by XmasNet and SLM

On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points that are not considered outliers. These results show that SLM outperforms XmasNet in classifying unseen PCa images. This suggests that SLM is able to adapt its architecture to the particular problem at hand, thus being able to outperform a state-of-the-art method for the challenging dataset at hand.

To statistically validate these results, a set of statistical tests was executed. In particular, as a first step, the Kolmogorov-Smirnov test was applied to test if the experimental results come from a normal distribution. The result of the test, performed with a significance level (α) of 0.05, suggested that the alternative hypothesis (i.e., data do not come from a normal distribution) cannot be rejected. Thus, a rank-based statistic was subsequently used. The Mann-Whitney U-test was performed with the null hypothesis that the samples have equal medians. As for the previous test, a significance level of 0.05 was considered. The p -value returned by the Mann-Whitney U-test was 0.016 thus suggesting that the difference (in terms of AUROC) among the considered techniques is statistically significant.

As previously discussed, it is fundamental to compare the time needed to perform the training process of the two networks. The results of this analysis are reported in Figure 5. From this comparison, it is appreciable the significant reduction (p -value returned

by the Mann-Whitney test smaller than 10^{-10}) in terms of the time introduced by the SLM.

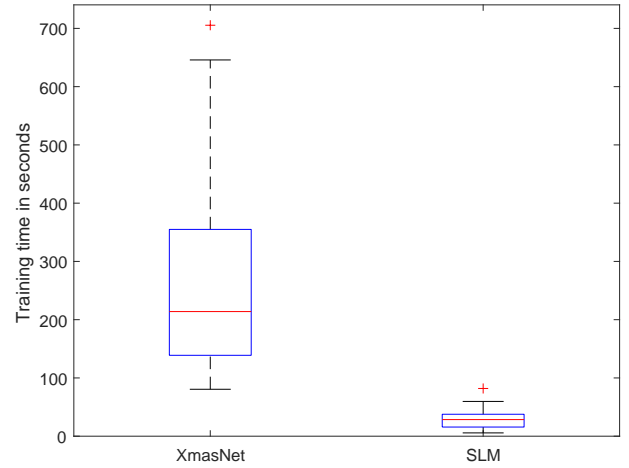


Figure 5: Training time required by XmasNet and SLM

In terms of speed-up (calculated as $\frac{\text{CNN}_{\text{time}}}{\text{SLM}_{\text{time}}}$), the average value obtained over the performed runs is 14.008. Thus, SLM outperforms XmasNet in terms of classification performance, and it is able to execute the training process faster than XmasNet by one order of magnitude. Additionally, it is important to emphasize that SLM is only run on CPU (whereas the XmasNet was trained using a GPU) and without any explicit parallelization. This further reinforces the results obtained, because each network evaluation would be suitably parallelized, thus achieving a higher speed-up. With regard to the SLM algorithm, Figure 6 conveys interesting information about the training process and the resulting network topology.

The SLM algorithm requires 12 iterations (median value) for producing the resulting network. This network has 2 hidden layers, approximately 50 hidden neurons and it is characterized by the presence of 40000 connections between the neurons. To conclude this section, it is important to strengthen the importance of the results achieved by the SLM *without* pre-training the CNN. This is a considerable advantage with respect to other existing models and techniques since the SLM is not bounded or dependent on backpropagation.

6 CONCLUSIONS

In this paper, multiparametric MRI images were the input of an image classification task proposed in the context of the PROSTATEx competition [28]. This challenge aimed at building a machine learning model for discriminating between benign and malignant PCa lesions. To tackle this compelling problem, the paper explored the use of the Semantic Learning Machine (SLM) neuroevolution algorithm to replace the backpropagation algorithm commonly used in the last fully-connected layers of CNNs.

To evaluate the suitability of the SLM algorithm, its performance was compared against the one achieved by XmasNet [29], a state-of-the-art CNN for addressing the PROSTATEx challenge [2]. Experimental results showed that SLM outperforms XmasNet both in

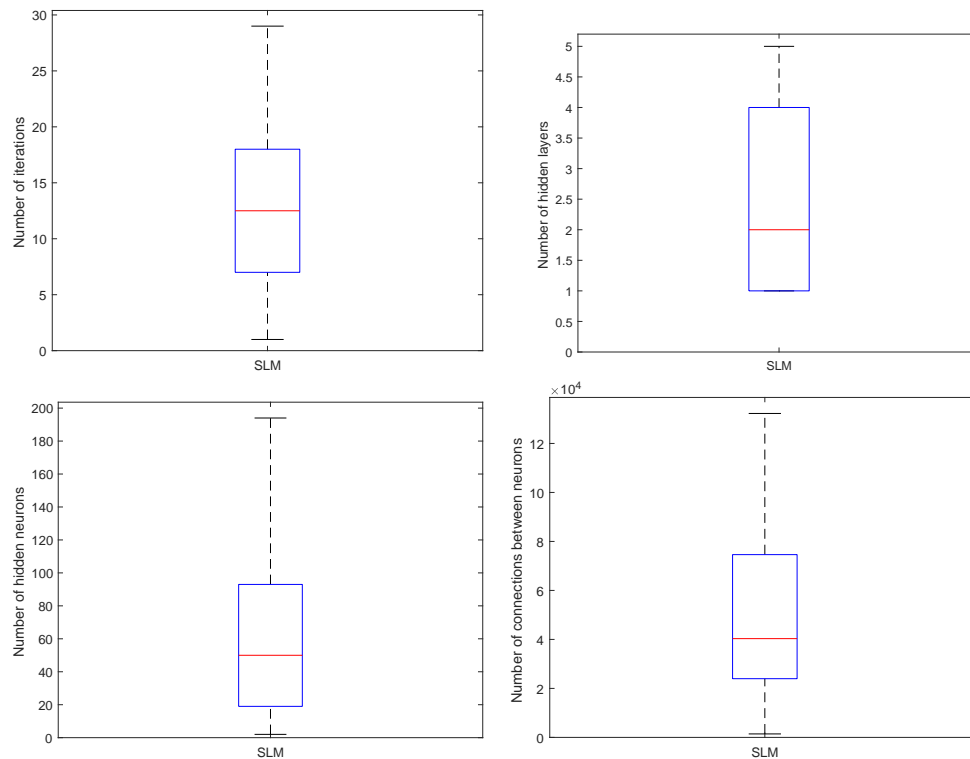


Figure 6: Information about the SLM training process: number of iterations, number of hidden layers, number of hidden neurons, and number of connections between neurons

terms of classification performance and running time (obtaining a speed-up of one order of magnitude).

While the results presented in this paper are promising and highlight the suitability of the SLM as a neuroevolution algorithm, this work can be extended in several directions. In particular, we are currently investigating the use of the SLM for optimizing the architecture of the convolutional layers of a CNN. This is the first step towards the most important objective of this research track, which is the definition of a novel SLM variant for optimizing the whole topology of a deep network.

Finally, the integration of Positron Emission Tomography (PET) imaging with multiparametric MRI [38] could further improve the PCa detection performance [26].

ACKNOWLEDGMENTS

This work was partially supported by projects UID/MULTI/00308/2019 and by the European Regional Development Fund through the COMPETE 2020 Programme, FCT - Portuguese Foundation for Science and Technology and Regional Operational Program of the Center Region (CENTRO2020) within project MAnAGER (POCI-01-0145-FEDER-028040). This work was also partially supported by national funds through FCT (Fundação para a Ciência e a Tecnologia) under project DSAIPA/DS/0022/2018 (GADgET).

REFERENCES

- [1] Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. <https://www.tensorflow.org/>. Software available from tensorflow.org.
- [2] Samuel G Armato, Henkjan Huisman, Karen Drukker, Lubomir Hadjiiski, Justin S Kirby, Nicholas Petrick, George Redmond, Maryellen L Giger, Kenny Cha, Artem Mamonov, et al. 2018. PROSTATEx Challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images. *J. Med. Imaging* 5, 4 (2018), 044501. <https://doi.org/10.1117/1.JMI.5.4.044501>
- [3] Léon Bottou. 2010. Large-scale machine learning with stochastic gradient descent. In *Proc. COMPSTAT'2010*. Springer, 177–186. https://doi.org/10.1007/978-3-7908-2604-3_16
- [4] Stefano Cagnoni, Andrew B Dobrzeniecki, Riccardo Poli, and Jacquelyn C Yanch. 1999. Genetic algorithm-based interactive segmentation of 3D medical images. *Image Vis. Comput.* 17, 12 (1999), 881–895. [https://doi.org/10.1016/S0262-8856\(98\)00166-8](https://doi.org/10.1016/S0262-8856(98)00166-8)
- [5] Young Jun Choi, Jeong Kon Kim, Namkug Kim, Kyoung Won Kim, Eugene K. Choi, and Kyoung-Sik Cho. 2007. Functional MR Imaging of Prostate Cancer. *RadioGraphics* 27, 1 (jan 2007), 63–75. <https://doi.org/10.1148/rg.271065078>
- [6] François Chollet et al. 2015. Keras. <https://keras.io>.
- [7] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. 2013. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging* 26, 6 (2013), 1045–1057. <https://doi.org/10.1007/s10278-013-9622-7>
- [8] Tyler Clark, Junjie Zhang, Sameer Baig, Alexander Wong, Masoom A Haider, and Farzad Khalvati. 2017. Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks. *J. Med. Imaging* 4, 4 (2017), 041307. <https://doi.org/10.1117/1.JMI.4.4.041307>
- [9] Anna Fabijańska. 2016. A novel approach for quantification of time-intensity curves in a DCE-MRI image series with an application to prostate cancer. *Comput. Biol. Med.* 73 (2016), 119–130. <https://doi.org/10.1016/j.combiomed.2016.04.010>
- [10] Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, and Ian Nimmo-Smith. 2014. DIPy, a library

- for the analysis of diffusion MRI data. *Front. Neuroinform.* 8 (2014), 8. <https://doi.org/10.3389/fninf.2014.00008>
- [11] Ivo Gonçalves, Sara Silva, and Carlos M Fonseca. 2015. On the generalization ability of geometric semantic genetic programming. In *Genetic Programming*. Springer, 41–52. https://doi.org/10.1007/978-3-319-16501-1_4
- [12] Ivo Gonçalves, Sara Silva, and Carlos M. Fonseca. 2015. Semantic learning machine: a feedforward neural network construction algorithm inspired by geometric semantic genetic programming. In *Progress in Artificial Intelligence*. Lecture Notes in Computer Science, Vol. 9273. Springer, 280–285. https://doi.org/10.1007/978-3-319-23485-4_28
- [13] Ivo Gonçalves, Sara Silva, Carlos M Fonseca, and Mauro Castelli. 2016. Arbitrarily close alignments in the error space: a geometric semantic genetic programming approach. In *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*. ACM, 99–100. <https://doi.org/10.1145/2908961.2908988>
- [14] Ivo Gonçalves, Sara Silva, Carlos M. Fonseca, and Mauro Castelli. 2017. Unsure when to Stop? Ask Your Semantic Neighbors. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*. ACM, New York, NY, USA, 929–936. <https://doi.org/10.1145/3071178.3071328>
- [15] Vikas Gulani, Fernando Calamante, Frank G Shellock, Emanuel Kanal, Scott B Reeder, et al. 2017. Gadolinium deposition in the brain: summary of evidence and recommendations. *Lancet Neurol.* 16, 7 (2017), 564–570. [https://doi.org/10.1016/S1474-4422\(17\)30158-8](https://doi.org/10.1016/S1474-4422(17)30158-8)
- [16] Yanrong Guo, Yaozong Gao, and Dinggang Shen. 2016. Deformable MR prostate segmentation via deep feature learning and sparse patch matching. *IEEE Trans. Med. Imaging* 35, 4 (2016), 1077–1089. <https://doi.org/10.1109/TMI.2015.2508280>
- [17] Masoom A Haider, Theodoros H Van Der Kwast, Jeff Tanguay, Andrew J Evans, Ali-Tahir Hashmi, Gina Lockwood, and John Trachtenberg. 2007. Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer. *AJR Am. J. Roentgenol.* 189, 2 (2007), 323–328. <https://doi.org/10.2214/AJR.07.2211>
- [18] Thomas Hambroek, Pieter C Vos, Christina A Hulsbergen-van de Kaa, Jelle O Barentsz, and Henkjan J Huisman. 2013. Prostate cancer: computer-aided diagnosis with multiparametric 3-T MR imaging—effect on observer performance. *Radiology* 266, 2 (2013), 521–530. <https://doi.org/10.1148/radiol.12111634>
- [19] Jan-Benedikt Jagusch, Ivo Gonçalves, and Mauro Castelli. 2018. Neuroevolution under unimodal error landscapes: an exploration of the semantic learning machine algorithm. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO) Companion*. ACM, New York, NY, USA, 159–160.
- [20] Haozhe Jia, Yong Xia, Yang Song, Weidong Cai, Michael Fulham, and David Dagan Feng. 2018. Atlas registration and ensemble deep convolutional neural network-based prostate segmentation using magnetic resonance imaging. *Neurocomputing* 275 (2018), 1358–1369. <https://doi.org/10.1016/j.neucom.2017.09.084>
- [21] Daniel Junker, Fabian Steinkohl, Veronika Fritz, Jasmin Bektic, Theodoros Tokas, Friedrich Aigner, Thomas RW Herrmann, Michael Rieger, and Udo Nagele. 2018. Comparison of multiparametric and biparametric MRI of the prostate: are gadolinium-based contrast agents needed for routine examinations? *World J. Urol.* (2018), 1–9. <https://doi.org/10.1007/s00345-018-2428-y>
- [22] Konstantinos Kamnitsas, Christian Ledig, V. F. J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, et al. 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* 36 (2017), 61–78. <https://doi.org/10.1016/j.media.2016.10.004>
- [23] Diederik P Kingma and Jimmy Ba. 2014. Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [24] Deukwoo Kwon, Isildinha M Reis, Adrian L Breto, Yohann Tschudi, Nicole Gautney, Olmo Zavala-Romero, Christopher Lopez, John C Ford, Sanoo Punnen, Alan Pollack, et al. 2018. Classification of suspicious lesions on prostate multiparametric MRI using machine learning. *J. Med. Imaging* 5, 3 (2018), 034502. <https://doi.org/10.1117/1.JMI.5.3.034502>
- [25] Guillaume Lemaître, Robert Martí, Jordi Freixenet, Joan C Vilanova, Paul M Walker, and Fabrice Meriaudeau. 2015. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. *Comput. Biol. Med.* 60 (2015), 8–31. <https://doi.org/10.1016/j.combiomed.2015.02.009>
- [26] Roger Li, Gregory C Ravizzini, Michael A Gorin, Tobias Maurer, Matthias Eiber, Matthew R Cooperberg, Mehrdad Alemozaffar, Matthew K Tollefson, Scott E Delacroix, and Brian F Chapin. 2018. The use of PET/CT in prostate cancer. *Prostate Cancer Prostatic Dis.* 21, 1 (2018), 4. <https://doi.org/10.1038/s41391-017-0007-8>
- [27] Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman. 2014. Computer-aided detection of prostate cancer in MRI. *IEEE Trans. Med. Imaging* 33, 5 (2014), 1083–1092. <https://doi.org/10.1109/TMI.2014.2303821>
- [28] Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman. 2017. “PROSTATEx Challenge data”, The Cancer Imaging Archive. <https://wiki.cancerimagingarchive.net/display/Public/SPIE-AAPM-NCI+PROSTATEx+Challenges>. Online; Accessed on January 25, 2019.
- [29] Saifeng Liu, Huaixiu Zheng, Yesu Feng, and Wei Li. 2017. Prostate cancer diagnosis using deep learning with 3D multiparametric MRI. In *Medical Imaging 2017: Computer-Aided Diagnosis (Proceedings SPIE)*, Vol. 10134. International Society for Optics and Photonics, 1013428. <https://doi.org/10.1117/1.2277121>
- [30] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. 1997. Multi-modality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging* 16, 2 (apr 1997), 187–198. <https://doi.org/10.1109/42.563664>
- [31] Lukasz Matulewicz, Jacobus FA Jansen, Louisa Bokacheva, Hebert Alberto Vargas, Oguz Akin, Samson W Fine, Amita Shukla-Dave, James A Eastham, Hedvig Hricak, Jason A Koutcher, et al. 2014. Anatomic segmentation improves prostate cancer detection with artificial neural networks analysis of 1H magnetic resonance spectroscopic imaging. *J. Magn. Reson. Imaging* 40, 6 (2014), 1414–1421. <https://doi.org/10.1002/jmri.24487>
- [32] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In *Proceedings of the International Conference on 3D Vision (3DV)*. IEEE, 565–571. <https://doi.org/10.1109/3DV.2016.79>
- [33] Alberto Moraglio, Krzysztof Krawiec, and Colin G Johnson. 2012. Geometric semantic genetic programming. In *Parallel Problem Solving from Nature-PPSN XII*. Springer, 21–31. https://doi.org/10.1007/978-3-642-32937-1_3
- [34] Alberto Moraglio and Andrea Mambrini. 2013. Runtime analysis of mutation-based geometric semantic genetic programming for basis functions regression. In *Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation (GECCO)*. ACM, 989–996. <https://doi.org/10.1145/2463372.2463492>
- [35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: convolutional Networks for Biomedical Image Segmentation. In *Proceedings of the Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (LNCS)*, Vol. 9351. Springer, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- [36] Andrew B Rosenkrantz, Ruth P Lim, Merhad Haghighi, Molly B Somberg, James S Babb, and Samir S Taneja. 2013. Comparison of interreader reproducibility of the prostate imaging reporting and data system and Likert scales for evaluation of multiparametric prostate MRI. *AJR Am. J. Roentgenol.* 201, 4 (2013), W612–W618. <https://doi.org/10.2214/AJR.12.10173>
- [37] Leonardo Rundo, Carmelo Militello, Giorgio Russo, Antonio Garufi, Salvatore Vitabile, Maria Carla Gilardi, and Giancarlo Mauri. 2017. Automated prostate gland segmentation based on an unsupervised fuzzy c-means clustering technique using multispectral T1w and T2w MR imaging. *Information* 8, 2 (2017), 49.
- [38] L. Rundo, A. Stefano, C. Militello, G. Russo, M. G. Sabini, C. D’Arrigo, F. Marletta, M. Ippolito, G. Mauri, S. Vitabile, and M. C. Gilardi. 2017. A fully automatic approach for multimodal PET and MR image segmentation in Gamma Knife treatment planning. *Comput. Methods Programs Biomed.* 144 (2017), 77–96. <https://doi.org/10.1016/j.cmpb.2017.03.011>
- [39] Leonardo Rundo, Andrea Tangherloni, Paolo Cazzaniga, Marco S Nobile, Giorgio Russo, Maria Carla Gilardi, Salvatore Vitabile, Giancarlo Mauri, Daniela Besozzi, and Carmelo Militello. 2019. A novel framework for MR image segmentation and quantification by using MedGA. *Comput. Methods Programs Biomed.* (2019). <https://doi.org/10.1016/j.cmpb.2019.04.016> In press.
- [40] Leonardo Rundo, Andrea Tangherloni, Carmelo Militello, Maria Carla Gilardi, and Giancarlo Mauri. 2016. Multimodal medical image registration using particle swarm optimization: a review. In *Proc. Symposium Series on Computational Intelligence (SSCI)*. IEEE, 1–8. <https://doi.org/10.1109/SSCI.2016.7850261>
- [41] Leonardo Rundo, Aandrea Tangherloni, Marco S. Nobile, Carmelo Militello, Daniela Besozzi, Giancarlo Mauri, and Paolo Cazzaniga. 2019. MedGA: a novel evolutionary method for image enhancement in medical imaging systems. *Expert Syst. Appl.* 119 (2019), 387–399. <https://doi.org/10.1016/j.eswa.2018.11.013>
- [42] Angela Serra, Paola Galdi, and Roberto Tagliaferri. 2018. Machine learning for bioinformatics and neuroimaging. *Wiley Interdisc. Rev. Data Min. Knowl. Discov.* 8, 5 (2018), e1248. <https://doi.org/10.1002/widm.1248>
- [43] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. 2019. Cancer statistics, 2019. *CA Cancer J. Clin.* 69, 1 (2019), 7–34. <https://doi.org/10.3322/caac.21551>
- [44] Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the International Conference on Learning Representations (ICLR)*. arXiv preprint arXiv:1409.1556.
- [45] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. 2014. Striving for simplicity: The all convolutional net. In *Proceedings of the International Conference on Learning Representations (ICLR)*. 1–14. arXiv preprint arXiv:1412.6806.
- [46] Radka Stoyanova, Mandeep Takhar, Yohann Tschudi, John C Ford, Gabriel Solórzano, Nicholas Erho, Yoganand Balagurunathan, Sanoo Punnen, Elai Davicioni, Robert J Gillies, et al. 2016. Prostate cancer radiomics and the promise of radiogenomics. *Transl. Cancer Res.* 5, 4 (2016), 432.
- [47] Yohan Sumathipala, Nathan Lay, Baris Turkbey, Clayton Smith, Peter L Choyke, and Ronald M Summers. 2018. Prostate cancer detection from multi-institution multiparametric MRIs using deep convolutional neural networks. *J. Med. Imaging* 5, 4 (2018), 044507. <https://doi.org/10.1117/1.JMI.5.3.034502>
- [48] Jinquan Sun, Yinghuan Shi, Yang Gao, and Dinggang Shen. 2017. A point says a lot: an interactive segmentation method for MR prostate via one-point labeling. In *Proceedings of the International Workshop on Machine Learning in Medical Imaging (MLMI)*. Springer, 220–228. https://doi.org/10.1007/978-3-319-67389-9_26
- [49] Baris Turkbey, Anna M Brown, Sandeep Sankineni, Bradford J Wood, Peter A Pinto, and Peter L Choyke. 2016. Multiparametric prostate magnetic resonance imaging in the evaluation of prostate cancer. *CA: A Cancer Journal for Clinicians*

- 66, 4 (2016), 326–336. <https://doi.org/10.3322/caac.21333>
- [50] Zhiwei Wang, Chaoyue Liu, Danpeng Cheng, Liang Wang, Xin Yang, and Kwang-Ting Cheng. 2018. Automated detection of clinically significant prostate cancer in mp-MRI images based on an end-to-end deep neural network. *IEEE Trans. Med. Imaging* 37, 5 (2018), 1127–1139. <https://doi.org/10.1109/TMI.2017.2789181>
- [51] David H Wolpert and William G Macready. 1997. No free lunch theorems for optimization. *IEEE Trans. Evol. Computat.* 1, 1 (1997), 67–82. <https://doi.org/10.1109/4235.585893>